Journal of Pharmacognosy and Phytochemistry
Available online at www.phytojournal.com

E-ISSN: 2278-4136
P-ISSN: 2349-8234
www.phytojournal.com
JPP 2020; 9(2): 826-830
Received: 13-01-2020
Accepted: 17-02-2020

Rahul Patil
Ph. D. Scholar, Department of Soil and Water Engineering, CAE, UAS, Raichur, Karnataka, India

Nagaraj DM
Ph. D. Scholar, Department of Soil and Water Engineering, CAE, UAS, Raichur, Karnataka, India

Shrikantagouda
M. Tech, Department of Soil and Water Engineering, CAE, UAS, Raichur, Karnataka, India

Santosha Rathod
Scientist, Indian Institute of Rice Research, ICAR, Hyderabad, Telangana, India

BS Polisgowdar
Professor Department of Irrigation and Drainage Engineering, CAE, UAS, Raichur, Karnataka, India

Corresponding Author:
Rahul Patil
Ph. D. Scholar, Department of Soil and Water Engineering, CAE, UAS, Raichur, Karnataka, India

Accuracy testing of box Jenkins models in forecasting potential evapotranspiration for Raichur district

Rahul Patil, Nagaraj DM, Shrikantagouda, Santosha Rathod and BS Polisgowdar

Abstract
The prediction of evapotranspiration is necessary for a reliable management of irrigation systems. This paper is based on models used for the prediction of potential evapotranspiration in the area of Raichur District, Karnataka, India. The potential evapotranspiration time series was estimated based upon Thornthwaite model and possible Seasonal ARIMA models were developed and best fit model were selected based on Least AIC and BIC values and forecasting was done for 1-6 months lead time. The obtained results show that the ARIMA model is a very effective and reliable prediction model for short term forecasts.

Keywords: forecasting potential evapotranspiration, irrigation systems, potential evapotranspiration

Introduction
Evapotranspiration (ET) is usually the largest component of the hydrologic cycle, given that most precipitation that falls on land is returned to the atmosphere. Globally, about 60% of the annual precipitation falling over the land surface is consumed by ET. Quantification of ET is used for many purposes, including crop production, water resources management, and environmental assessment (Aruna *et al*., 2017) [2]. It is important component for field water balance and needs to be accurately quantified. The amount of water supplied to meet the evapotranspiration requirements of agricultural crops dictates the quality and quantity of production in an area. The ET data for agricultural crops has become increasingly important in irrigation as well as in water resources management. The ET process is controlled by factors such as temperature, solar radiation, and humidity, which vary temporally and spatially (Mohan and Arumugam, 1995) [8].
The stochastic models are based on the time dependent variation and consider random effects involved in the ET process. Stochastic linear models are fitted to hydrological data or time series such as evapotranspiration series for two main reasons: it enables the integration of an on-farm system with the main system and it facilitates the real-time operation of an irrigation system. The synthetic and forecast data are of considerable importance to the design and operation of water resource systems. The most popular time series model is the autoregressive integrated moving average (ARIMA) model. In this model the forecast of a variable is defined as a linear combination of the previous state of variable and previous forecast error. The ARIMA process is a powerful time series modeling and forecasting technique which possesses flexibility for the inclusion of many time series characteristic. In past ARIMA models have been used successfully to hydrological time series model (Popale and Gorantiwar, 2014) [10]. Mohan and Arumugam (1995) [8] studied on seasonal ARIMA modelling of weekly data of evapotranspiration of annamlainagar meteorological station, India. Popale and Gorantiwar (2014) [10] used ARIMA model for forecasting rainfall of rahuri region, India. Gorantiwar and Patil (2009) [5] did analysis of ET_o of Rahuri region, India. Hamdi *et al*. (2008) [7] developed seasonal ARIMA model for the Jordan valley. Asadi *et al.* (2014) forecasts evapotranspiration for humid and semi-humid region. salas *et al*. (1980) [11] discussed in detail about time series modelling. The objective of this study is to establish a time series model to analyse and forecast reference crop evapotranspiration for the Raichur district.

Materials and Methodology
Raichur is an administrative district in the Indian state of karnataka. It is located in the northeastern part of the state and is bounded by Yadgir district in the north in the north, Bijou and Bagalkot district in the northwest, Koppel district in the west, Bellary district in the south,

Mahabubnagar district of Telangana and Kurnool district of Andhra Pradesh in the east. The district is bounded by the Krishna River on the north and the Tungabhadra River on the south. The wedge of land between the rivers is known as the Raichur Doab, after the city of Richer. Bijou and Yadira districts lie to the north across the Krishna River. Bagalkot and Koppel districts lie to the west. Across the Tungabhadra lies Bellary District of Karnataka to the southwest and Mahabubnagar of Telangana to the southeast. Kurnool District of Andhra Pradesh state lies to the east, and includes the lower portion of the Raichur Doab.

Thornth waite method (Potential evapotranspiration)
The potential evapotranspiration is calculated by:

$$PET = 16K\left(\frac{10T}{I}\right)^m$$

Where
T is monthly mean temperature (°C); I is heat index calculated as the sum of 12 month index values; m is the coefficient dependent on *I*.

$$m=6.75 \times 10^{-7} \cdot I^3 - 7.71 \times 10^{-7} \cdot I^2 + 1.79 \times 10^{-2} \cdot I + 0.492$$

K is a correction coefficient computed as a function of the latitude and month.

Description of the stochastic models
The stochastic models, which are often known as time series models have been used in scientific, economic and engineering applications for the analysis of time series. Time series modeling techniques have been shown to provide a systematic empirical method for simulating and forecasting the behavior of uncertain hydrologic systems and for quantifying the expected accuracy of the forecasts (Mishra and Desai, 2005) [9].

ARIMA models
Autoregressive (AR) models can be effectively coupled with moving average (MA) models to produce a general and useful class of time series models named autoregressive moving average (ARMA) models. In an ARMA model the current value of the time series is expressed as a linear aggregate of *p* previous values and a weighted sum of *q* previous deviations (original value minus fitted value of previous data) plus a random parameter.
However, an ARIMA model can be used when the data are stationary. This class of models can be extended to non-stationary series by allowing differencing of data series. These models are called autoregressive integrated moving average (ARIMA) models. Box and Jenkins (1976) [3] provides a new generation of forecasting tools, known as the ARIMA methodology, which emphasizes on analyzing the stochastic properties of time series on their own rather than constructing single or simultaneous equation models. ARIMA models allow each variable to be stated by its own lagged values and stochastic error terms. The general non-seasonal ARIMA model is AR to order *p* and MA to order *q* and operates on d^{th} difference of the time series z_t; thus a model of the ARIMA family is classified by three parameters (p, d, q) that can have zero or positive integral values (Mishra and Desai, 2005) [9]

The general non-seasonal ARIMA model may be written as

$$\emptyset(B)\nabla^d_{Z_t} = \theta(B)a_t$$

Where
θ (B) are polynomials of order p and q, respectively. Non-seasonal AR operator of order p is written as

$\emptyset(B) = \left(1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots \emptyset_p B^p\right)$ and non-seasonal MA operator of order q is written as

$$\theta(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots \theta_q B^q\right)$$

Seasonal ARIMA models
Many time series contain cyclic features. Very often in hydrologic time series these features are of an annual cycle primarily due to the earth's rotation about the sun. Such series are cyclically non-stationary. Once the deterministic cyclic effects have been removed from a series, the ARIMA approach can be applied to obtain a linear model for the stochastic part of the series. Box *et al*. (1994) [4] have generalized the ARIMA model to deal with seasonality, and define a general multiplicative seasonal ARIMA model, which are commonly known as SARIMA models. An inherent advantage of the SARIMA family of models is that few model parameters are required for describing time series, which exhibit non-stationary both within and across the seasons. In short notation the SARIMA model described as ARIMA *(p, d, q) (P, D, Q)s*, where *(p, d, q)* is the non-seasonal part of the model and *(P, D, Q)s* is the seasonal part of the model, which is mentioned below

$$\emptyset_p(B)\Phi_p(B^s)\nabla^d\nabla^D_s Z_t = \theta_p(B)\varphi_Q(B^s)a_t$$

Where
p is the order of non-seasonal auto regression, *d* the number of regular differencing, *q* the order of no seasonal MA, *P* the order of seasonal auto regression, *D* the number of seasonal differencing, *Q* the order of seasonal MA, *s* is the length of season, seasonal AR parameter of order P, seasonal MA parameter of order Q.

Model identification
This step consists of identifying the possible ARIMA model that represents the behavior of the time series. The series behavior was investigated by the autocorrelation function (ACF) and partial autocorrelation function (PACF). The ACF and PACF were used to assist in determining the order of the model. The information provided by ACF and PACF is useful to suggest the type of models that can be built. The final model was then selected using the Akaike information criterion (AIC) and Schwarz-Bayesian criterion (SBC). These criteria help to rank models (models having the lowest value of criterion being the best). The AIC and SBC take the mathematical form as shown below.

$$AIC= -2 \log (L) + 2k$$
$$SBC= -2 \log (L) + k \ln (n)$$

Where
k is number of parameters in the model, L is the likelihood function of the ARIMA model; and n is the number of observations.

Parameter estimation
After identifying the appropriate model as an essential step, the estimation of model parameters was achieved. The model

estimate values for the AR and MA parts were calculated using Maximum likelihood. The AR and MA parameters were tested to make sure that they are statistically significant or not. The associated parameters, such as standard error of estimates and their related t-values, are also calculated

Diagnostic checking
Diagnosing the ARIMA model is a crucial part and the last step of the model development. It involves in checking the adequacy of selected model. Several diagnostic statistics and plots of residuals are investigated to see if the residuals are correlated white noise or not. In this study the residual ACF function (RACF) was obtained to determine whether residuals are white noise.

Drought forecasting
The prediction of Potential evapotranspiration was done for 1-6 month lead time using the best fit models from historical data. Basic statistical properties of the observed and predicted data for 1-6 month lead time was computed and tested whether the predicted data preserve the basic statistical properties of the observed PET series. The predictions are calculated for different lead time. For instance, a 1-month lead time prediction means that during January 2017, the prediction for February 2017 is computed. The correlation coefficients (R), RMSE and MAE were observed between the observed and predicted data for 1 to 6 month lead times.

Results and Discussion
Development of model was done with certain prerequisite tests namely Stationarity and autocorrelation test. The autocorrelation test was carried out using box. Test and corresponding probability levels are presented in Table 1. The results reveals that the test statistic for box. Test with a Chi square and P values were 233.73 (0.01), 261.11 (0.01), 237.33 (0.01), 235.86 (0.01) and 235.86 (0.01) for Raichur, Manvi, Sindhanuru and Lingasugur respectively, were observed to be significant at 5 % level of significance, hence it can be understood that autocorrelation exists in data. On the other hand adf.test was carried out to check whether the data is stationary or not. The results reveal that data is observed to have a seasonality so seasonal differencing was done to the data sets Table 2.
The principal step in Box-Jenkins ARIMA model building is identification of the model. Different orders of Autoregressive (AR) and Moving Average (MA) parameters *p* and *q* are considered and combination of the order which yields maximum log-likelihood and lowest values of Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) are considered as best model. The results pertaining to Raichur, Manvi, Sindhanuru and Lingasuguru stations regarding model development are presented in Table 3 & 4. The ACF and PACF were plotted (Fig. 1 and 2) to determine the model, the data were observed have a seasonality in the data so seasonal ARIMA models were selected with a seasonal differencing as shown in Table 4. The best selected models for different stations were ARIMA $(2,0,2)(1,1,2)_{12}$, ARIMA $(1,0,1)(2,1,0)_{12,}$ ARIMA $(1,0,1)(1,1,2)_{12}$ and ARIMA $(1,0,1)(2,1,0)_{12}$ with an maximum likelihood values of -1380.54, -1205.10, -1338.08 and -1356.04 respectively for Raichur, Manvi, Sindhanuru and Lingasuguru. The parameters estimated for different stations are presented in Table 4. In addition, the residuals were obtained by differencing original series with the fitted series and residuals were found to be white noise as presented in Table 5.
Soon After the development of models for 4 taluks the forecasting part was carried out at different lead time (1-6 months) and the results Table 6 reveal that initially for all stations the forecast was observed to be good at 1 lead time with a correlation coefficient of 0.90, 0.93, 0.87 and 0.90 for Raichur, Manvi, Sindhanur and Lingasugur respectively. The RMSE and MAE were observed to be least at 1 leads and increases as the lead time increase, these stochastic models were found to suitable to forecast up to 1 lead time. A view at the Table 6 can be easily noticed that as the lead time increases the error rate has been increase tremendously. It can be easily concluded that Seasonal ARIMA models suits well for forecasting at 1 month lead time for Potential evapotranspiration forecasting under Raichur Region. Basic statistical properties are compared between observed and forecasted data for one month lead time, using t-test for the means and F-test for standard deviation (Haan 1977) [6], shown in Table 7. Since t_{cal} values related to means were between t-critical table values (±1.71 for two tailed at a 5% significance level), the data shows that there is no significant difference between the mean values of observed and predicted data. Similarly, the F_{cal} values of standard deviation were smaller than the F-critical values at a 5% significance level. Thus, the results show that predicted data preserves the basic statistical properties of the observed series.

Table 1: Auto correlation test for different station

Station	Chi-Square	Lag order	P-value
Raichur	233.73	1	<0.001
Manvi	261.11	1	<0.001
Sindanuru	237.33	1	<0.001
Lingasuguru	235.86	1	<0.001

Table 2: Stationarity test for different station

Station	Dickey fuller	Lag order	P-value
Raichur	-19.691	7	0.01
Manvi	-18.671	7	0.01
Sindanuru	-19.649	7	0.01
Lingasuguru	-19.841	7	0.01

Table 3: Log likelihood AIC and BIC values of ARIMA model for different station

Stations	Model	Log-Likelihood	AIC	BIC
Raichur	ARIMA $(2,0,2)(1,1,2)_{12}$	-1380.54	2777.07	2808.93
Manvi	ARIMA $(1,0,1)(2,1,0)_{12}$	-1205.10	2422.01	2445.9
Sindhanuru	ARIMA $(1,0,1)(1,1,2)_{12}$	-1338.08	2688.16	2712.05
Lingasuguru	ARIMA $(1,0,1)(2,1,0)_{12}$	-1356.04	2724.08	2747.97

Table 4: Parameter estimation of ARIMA by maximum likelihood method for different station

Station	Model	Parameters	Estimate	S.E.	Z value	P-value
Raichur	ARIMA $(2,0,2)(1,1,2)_{12}$	AR1	-0.262	0.144	-1.811	0.069
		AR2	0.605	0.117	5.165	< 0.001
		MA1	0.560	0.169	3.324	< 0.001
		MA2	-0.389	0.157	-2.477	0.013
		SAR1	0.256	0.323	0.791	0.429
		SMA1	-1.268	0.310	-4.096	< 0.001
		SMA2	0.360	0.285	1.264	0.206
Manvi	ARIMA $(1,0,1)(2,1,0)_{12}$	AR1	0.636	0.111	5.736	< 0.001
		MA1	-0.279	0.143	-1.949	0.05
		SAR1	-0.747	0.047	-15.787	< 0.001
		SAR2	-0.373	0.048	-7.793	< 0.001
Sindhanuru	ARIMA $(1,0,1)(1,1,2)_{12}$	AR1	0.669	0.114	5.858	< 0.001
		MA1	-0.370	0.146	-2.533	0.01
		SAR1	0.101	0.308	0.327	0.74
		SMA1	-1.165	0.300	-3.887	< 0.001
		SMA2	0.264	0.285	0.863	0.38
Lingasuguru	ARIMA $(1,0,1)(2,1,0)_{12}$	AR1	0.624	0.122	5.109	< 0.001
		MA1	-0.316	0.152	-2.083	0.03
		SAR1	0.122	0.309	0.396	0.69
		SMA1	-1.177	0.299	-3.935	< 0.001
		SMA2	0.257	0.285	0.903	0.36

Table 5: Auto correlation check for residuals of ARIMA model at different station

Station	Chi-Square	Lag order	P-value
Raichur	2.10	1	0.14
Manvi	0.04	1	0.82
Sindhanuru	0.01	1	0.91
Lingasugur	2.27	1	0.13

Table 6: Performance measure of Seasonal ARIMA models at different stations

Station	Model	Performance measures	Lead time					
			1	2	3	4	5	6
Raichur	ARIMA $(2,0,2)(1,1,2)_{12}$	RMSE	17.43	24.80	25.29	24.84	23.14	22.15
		MAPE	23.31	14.56	14.25	13.75	15.14	13.93
		MAE	13.64	21.10	20.55	19.81	20.10	17.35
		R	0.90	0.83	0.85	0.83	0.80	0.80
Manvi	ARIMA $(1,0,1)(2,1,0)_{12}$	RMSE	10.71	17.99	19.62	18.49	17.53	17.40
		MAPE	9.27	16.56	18.10	17.66	17.31	16.86
		MAE	8.99	15.92	16.68	15.95	15.13	13.86
		R	0.93	0.83	0.84	0.84	0.81	0.80
Sindhanuru	ARIMA $(1,0,1)(1,1,2)_{12}$	RMSE	10.52	20.22	39.54	52.15	59.60	61.46
		MAPE	5.28	10.55	22.50	29.21	33.49	37.85
		MAE	7.70	15.77	32.81	43.15	48.93	51.73
		R	0.87	0.77	0.70	0.71	0.65	0.54
Lingasuguru	ARIMA $(1,0,1)(2,1,0)_{12}$	RMSE	20.83	23.05	53.72	61.98	64.58	66.71
		MAPE	10.51	40.34	30.67	35.9	41.02	45.01
		MAE	15.59	33.54	44.59	51.29	54.88	56.32
		R	0.90	0.81	0.82	0.81	0.79	0.77

Table 7: Comparison of statistic properties of the observed and predicted data

Stations	Mean observed	Mean forecasted	Decision (t<1.71)	Observed variance	Forecast variance	Decision (f < 4.05)
Raichur	137.86	134.19	1.03	1392.85	994.21	0.14
Manvi	97.75	97.33	0.18	640.59	547.634	0.0008
Sindhanuru	136.74	136.54	0.009	1088.28	1397.82	0.0003
Lingasuguru	136.64	136.96	-0.072	1257.64	1546.41	0.0008

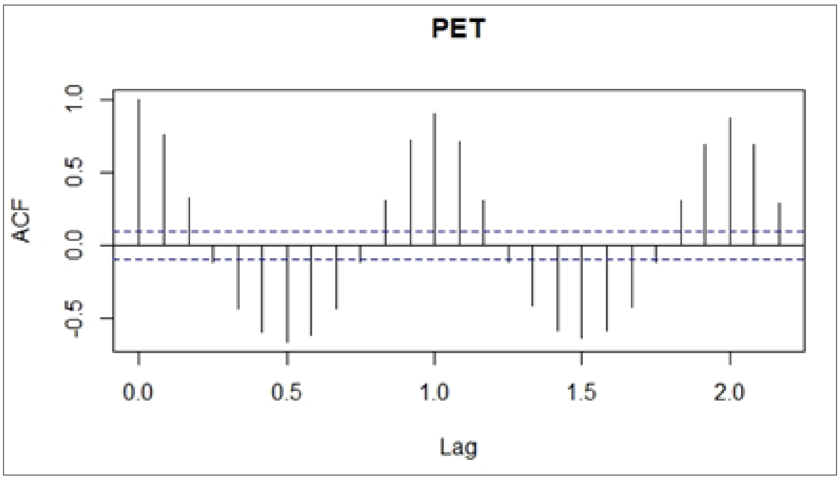

Fig 1: Autocorrelation function plot for Raichur Station

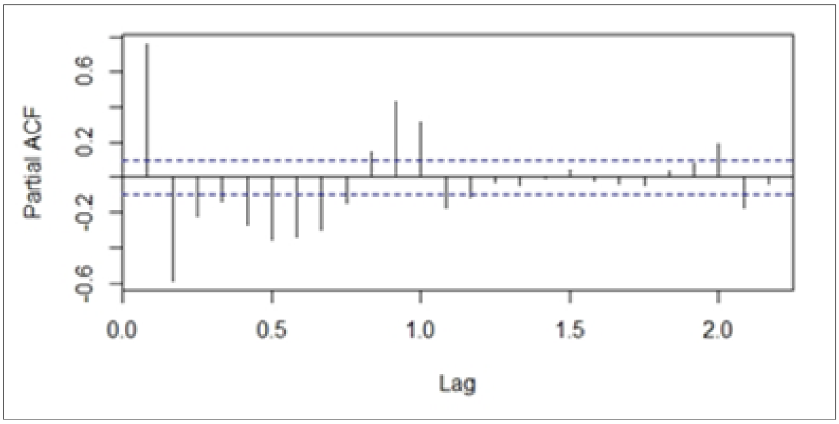

Fig 2: Partial autocorrelation function plot for Raichur Station

Conclusion

A view over the results predicted by the Seasonal ARIMA models reveals that the models have an ability to forecast up to 1 one month lead month with a higher accuracy over all the stations. Of the all stations, Seasonal ARIMA model provided excellent results at Sindhanuru station with an MAE, RMSE and MAPE values of 7.70, 10.52 and 5.28 respectively. Similarly for the basic statistical analysis the difference between the observed and forecasted mean were found to be no significant. The prediction of evapotranspiration guarantees reliable project planning, design and operating of irrigation systems.

References

1. Asadi A, Vahdat SF, Sarraf A. The forecasting of potential evapotranspiration using time series analysis in humid and semi humid regions, American Journal of Engineering Research. 2013; 2(1):296-302.
2. Aruna KT, Satish Kumar U, Ayyanagowdar MS, Srinivasa Reddy GV, Shanwad UK. Estimation of crop coefficient (Kc) values for groundnut crop with evaluated crop evapotranspiration at different moisture deficit levels under agro climatic condition of Raichur, Karnataka", *Green Farming.* 2017; 8(5):1161-1164.
3. Box GEP, Jenkins GM, Time series analysis forecasting and control, Holden-Day, San Francisco, 1976.
4. Box GEP, Jenkins GM, Reinsel GC, Time Series Analysis, Forecasting and Control, (3rd Edition), Englewood Cliffs NJ. Prentice Hall, 1994.
5. Gorantiwar SD, Patil PD. Stochastic modelling for crop evapotranspiration for Rahuri Region, *International* Journal of Agricultural Engineering. 2009; 2(1):140-145.
6. Haan CT. Statistical methods in hydrology, Iowa State Press, Iowa, 1977.
7. Hamdi MR, Bdour AN, Tarawneh AN. Developing reference crop evapotranspiration time series simulation model using class a pan: A case study for the Jordan Valley, Jordan, Journal of Earth and Environmental Sciences. 2008; 1(1):33-44.
8. Mohan S, Arumugam N. Forecasting weekly reference crop evapotranspiration series. Hydrological Sciences Journal. 1995; 40(6):689-702.
9. Mishra AK, Desai VR. Drought forecasting using stochastic models J Stoch. Environ. Res. Risk Assess. 2005; 19(1):326-339.
10. Popale PG, Gorantiwar SD. Stochastic generation and forecasting of weekly rainfall for Rahuri Region, International Journal of Innovative Research in Science, Engineering and Technology. 2014; 3(4):185-196.
11. Salas JD, Delleure JW, Yevjevich V, Lane WL. Applied Modeling of Hydrologic Time Series", Water Res. Publ., Colorado, USA, 1980.